Volume 13 Number 18 1985 Nucleic Acids Research

Detection of a unique human VκIV germline gene by a cloned cDNA probe

Philip Marsh, Frederick Mills and Hannah Gould

Department of Biophysics, King's College, 26-29 Drury Lane, London WC2B 5RL, UK

Received 30 July 1985; Accepted 6 September 1985

ABSTRACT

We have cloned the cDNA encoding the κIV chain of a human antibody with specificity against the major carbohydrate antigen of Streptococcus A. The cDNA has been used as a genetic probe to estimate the number of germline VκIV genes in human DNA. The presence of unique hybridizing bands on digestion of human DNA with several restriction endonucleases and the equivalence of the DNA in a band to a single gene per haploid genome point to the conclusion that there is a unique human VκIV germline gene. The corollary of this conclusion is that the diversity of human VκIV chains must be exclusively due to somatic mutation. This is supported by examination of the sequences of human κIV chain genes and their κIV chain products. Fusion of the unique germline VκIV gene (1) with one of several Jκ segments, followed by somatic mutations in the V region of the rearranged κIV gene, can account for the known sequences. The restricted germline gene repertoire may account for the small proportion of human κIV chains in the human κ chain sequence library (2).

INTRODUCTION

Human antibody κ chains have been classified into four subgroups on the basis of invariant amino acid sequences (3). Subgroup IV, which was originally described on the basis of sequencing of a small group of cryoglobulins (4), constitutes altogether only about 5% of the human chains of known sequence (2). The poor representation of subgroup IV chains can now be explained by the restricted germline gene repertoire. We present here two lines of evidence that human κIV chains are encoded by the products of somatic mutation of a unique VκIV germline gene. We first present evidence from hybridization studies with a subgroup IV-specific cDNA probe for the existence of a unique human germline VκIV gene. We then consider the available DNA and protein sequence information in view of the mechanisms of somatic mutation that are widely presumed to occur in immunoglobulin genes.

Our estimation of the VκIV gene repertoire involved the cloning of a VκIV cDNA encoding the light chain of an IgM(κ) antibody with

specificity against the immunodominant group of the Streptococcus A carbohydrate antigen, N-acetylglucosamine. The antibody is one of two produced by the biclonal lymphoblastoid cell line B17X2, which was established from human lymphocytes selected on the basis of affinity for the Streptococcus A antigen (5,6). Messenger RNA was isolated from B17X2 cells grown in tissue culture and a fraction enriched in κ chain mRNA was obtained by sucrose gradient centrifugation (7). The mRNA was transcribed into cDNA and recombined with a plasmid vector. The DNA was used to transform bacteria and colonies were screened using a human κ cDNA probe (8).

The total number of human Vκ genes may be calculated from the results of two previous hybridization studies to be around 20-30 genes. The genes corresponding to subgroups I, III and IV were estimated in one study with Vκ I probes to be 15 to 20 in number (9). The genes corresponding to the three individual subgroups were not resolved in this study because the Vκ I gene probes contained sequences that crosshybridize with VκIII and VκIV genes (9). In the study of subgroup II, a subgroup-specific probe revealed the existence of about 10-15 more human Vκ genes (10), bringing the total to 20-30. The collection of cloned and sequenced human Vκ genes, excluding the obvious pseudogenes (11,1), adds up to about the same number.

We have analyzed the gene family corresponding to subgroup IV using a subgroup specific probe. Since the sequence encoding the third framework region (FR3) is highly conserved and causes crosshybridization of Vκ I, Vκ III and Vκ IV(9), we made a probe lacking this sequence. This was obtained by subcloning an appropriate restriction fragment from the cloned B17 κIV chain cDNA. We have used the subgroup-specific probe in two hybridization assays to determine the number of germline Vκ IV genes. We first analyzed the restriction digests of human DNA by probing Southern blots with the subgroup-specific probe. Each of the digests contained a single hybridizing band. We next measured the absolute quantity of the hybridizing sequence in a single band. This was done by densitometry of the autoradiograph from a Southern blot containing a known amount of nuclease digested total human DNA. A calibration curve of the density vs. sequence concentration was constructed from analysis of a similar band in parallel lanes containing varying known amounts of the cloned gene sequence. This led to an estimate of 0.5 Vκ IV genes per haploid genome in human DNA.

The hybridization results eliminate the possibility that several genes may contribute to the diversity of human VκIV chains. The corollary of this conclusion is that the diversity of immunoglobulins in this subgroup can be generated only by somatic processes. We argue that the DNA and protein sequences differences within subgroup IV are within the bounds of the currently accepted norms for the somatic diversification of immunoglobulin genes.

METHODS

Cell Culture and Isolation of mRNA

B17X2 cells were grown as described by Polke *et al* (6). RNA was isolated from cells by guanidine thiocycanate (7) and the mRNA was purified by oligo-dT chromatography (7). The mRNA was fractionated by sucrose gradient centrifugation and assayed by *in vitro* translation, followed by immunoprecipitation and electrophoresis of the product (7).

cDNA Cloning

Double-stranded cDNA was prepared by the method of Land et al (12) with the following modifications: (1) Monovalent cation and actinomycin D were omitted from the oligo-dT primed first strand synthesis; (2) The dC tailing reactions contained 100 mM HEPES, pH 7.0, instead of Tris cacodylate; (3) the dCTP concentration was 0.05 mM in the cDNA tailing reactions; (4) the double-stranded cDNA was fractionated on a Bio-gel A150M agarose column (13).

The plasmid pAT153 (14) was prepared for recombination with the cDNA by cleavage with Pst I and isolation of the linear DNA by electrophoresis through agarose. The linear DNA was recovered from agarose by electroelution and purified by DEAE cellulose chromatography (15). The dG tails were added in a reaction mixture containing 2.5 µg pAT153, 100 mM HEPES, pH 7.0, 0.1mM dithiothreitol, 4 mM $MgCl_2$, 200µm dGTP, including 40 µCi ^{3}HdGTP (Amersham) and 5 units of terminal transferase (PL Biochemicals). Vector and cDNA were annealed at an estimated molar ratio of 1:1 in 200 mM NaCl, 10 mM Tris HCl, pH 8.0, 1 mM EDTA. The annealing mixture contained 10 ng dG-tailed pAT153 in 10 µl, which was overlaid with paraffin oil, heated to 68°C in a large water bath and allowed to cool over several hours.

Transformation of E. coli DH1 cells (16) was performed as described (16). Transformants were picked onto nitrocellulose filters in ordered arrays and grown up overnight on agar plates supplemented with 12.5

μg/ml tetracycline. Colony hybridization was carried out as described by Grunstein and Hogness (17). The DNA probe was labelled by nick-translation (18).

DNA Sequencing

DNA sequencing was carried out according to Maxam and Gilbert (19) with the following changes. Formic acid was used instead of pyridinium formate to modify A and G residues (20). DNA fragments were labelled at the 3'-ends by ^{32}p dCTP or ^{32}p GTP using T4 DNA polymerase (PL Biochemicals)(21).

Hybridization Analysis

After digestion of human DNA with restriction endonucleases, the fragments were separated by electrophoresis and transferred to nitrocellulose paper (20), hybridized to probes labelled by nick-translation (18) at low or high stringency (9), and analyzed by autoradiography and densitometry using routine methods. For the gene titrations a VκIV probe was made by self-ligating the subcloned fragment to produce multimers approximately 3 kb in length. This probe gave a significantly greater signal than the monomeric probe.

RESULTS

The screening of our B17X2 cDNA clone bank yielded two sets of clones, corresponding to the κ chains of the two antibodies produced by the mixed cell line (5,6). The cDNA inserts were excised from the plasmid vectors by digestion with Pst 1 and the sizes determined from their mobility on gel electorphoresis compared to known molecular weight markers. The longest cDNA inserts were further characterized by DNA sequencing to reveal the identity of the encoded κ chains. One of the cDNA sequences of 900 base pairs encodes a chain of subgroup IV and is designated pB17κIV.

Fig. 1 shows the restriction map of pB17κIV and the DNA sequencing strategy used to obtain the V region sequence in Fig. 2. The amino acid sequence shown in Fig. 2 was deduced from the genetic code. This sequence indicates that the chain encoded by pB17κIV is a member of subgroup IV. It is identical to that of the myeloma LEN (2) at 95 out of 101 residues (See Fig. 7). We note the extended form of the first complementarity determining region, which contains six amino acids more than all but two of the sequenced myeloma chains, viz LEN (subgroup IV) and CUM (subgroup II). The extra amino acids, 27A-F, are between

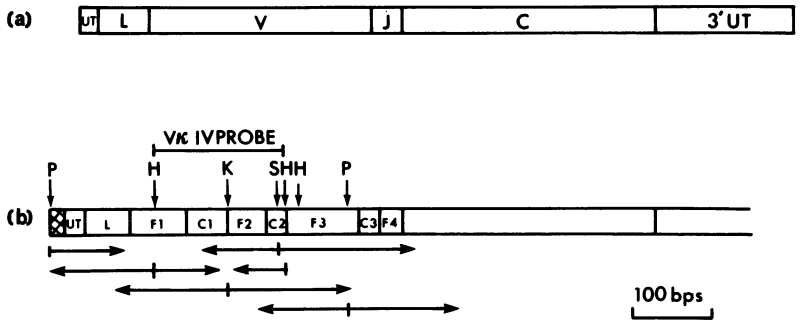

Figure 1. (a) Schematic diagram of a mRNA showing the variable (V), joining (J) and constant (C) region coding sequences and the 5' and 3' untranslated sequences. (b) the pBl7 IV cDNA restriction map and DNA sequencing strategy. The cutting sites of the following restriction endonucleases are indicated: P = Pst I, H = HinF I, K = Kpn I and S = Sma I. The regions of the cDNA corresponding to the 5' untranslated sequence of the mRNA (UT), and the coding sequences for the leader peptide (L), the four framework regions (Fl, F2, F3 and F4) and three complementarity determining regions (Cl, C2 and C3) are shown. Crosshatching shows G:C tails introduced by the cloning method.

the positions usually designated 27 and 28(2). The cloned cDNA contains the entire mature protein coding sequence, as well as that encoding the 20 amino acid leader peptide, and 25 base pairs corresponding to the 5' untranslated sequence of the mRNA. These features are included in the map in Fig. 1. The VκIV sequence in pBl7κIV is fused to Jκ1 without sequence alteration from the germline sequence (22).

To allow comparison with earlier work (9), we analyzed the digests

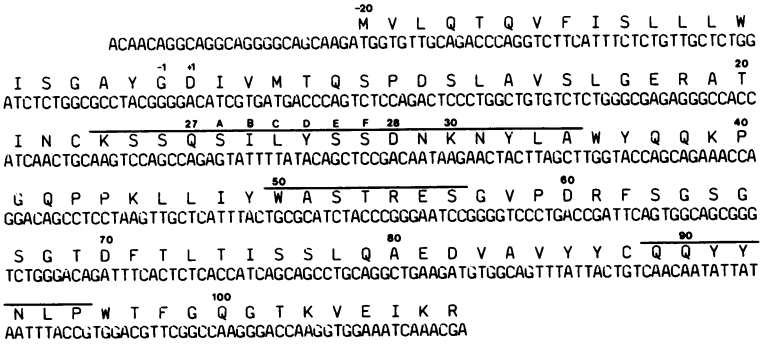

Figure 2. V region DNA sequence of the cDNA in pBl7 IV and the amino acid sequence of its translation product. The three complementarity determining regions are overlined.

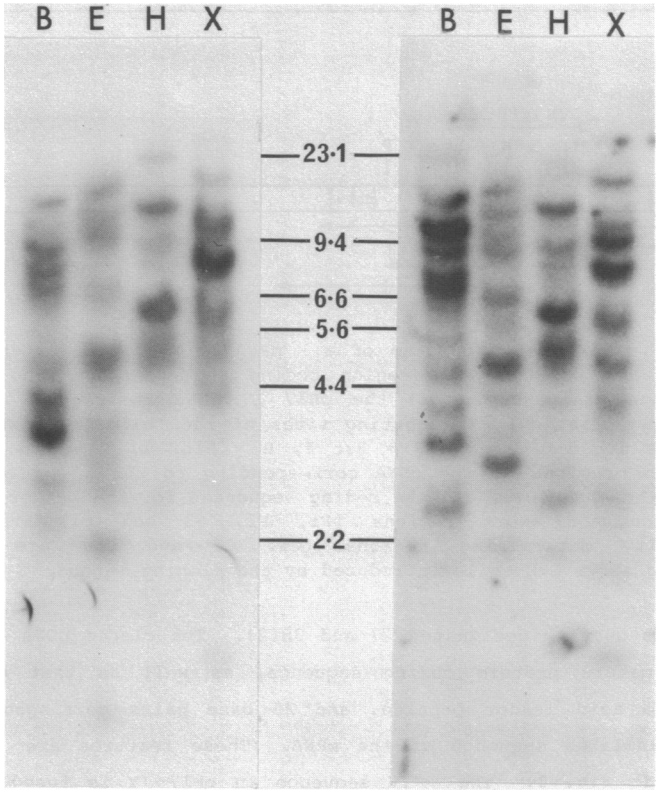

Figure 3. Restriction Fragments in Human Myeloid Cell DNA that hybridize to Human K Gene Probes at Low Stringency. DNA was digested to completion with Bgl II (B), EcoR I (E), Hind III (H) and Xba I (X), subjected to electrophoresis on an agarose gel, transferred to nitrocellulose and hybridized to either of two labelled probes at low stringency: autoradiograms corresponding to (left) the V 1 probe HK101/80(9) and (right) the pB17KIV probe. The positions of size markers obtained by the digestion of PM2 and lambda phage DNA with Hind III are indicated.

of human DNA with a VκI probe, as well as with our cDNA probe and the subgroup-specific probe derived from the cDNA. Human DNA was isolated from the myeloid cell line K562 (23) and digested with four restriction endonucleases, Bgl II, EcoR I, Hind III and Xba I. The results of the analysis of these digests with the VκI and cDNA probes are shown in Figs. 3 and 4. Under low stringency hybridization conditions (2 x SSC, 65°C), both probes detect, at varying intensities, many common bands in a complex pattern, as seen in Fig. 3. In addition, the cDNA probe detects a constant region (CK) fragment in each digest, which appears

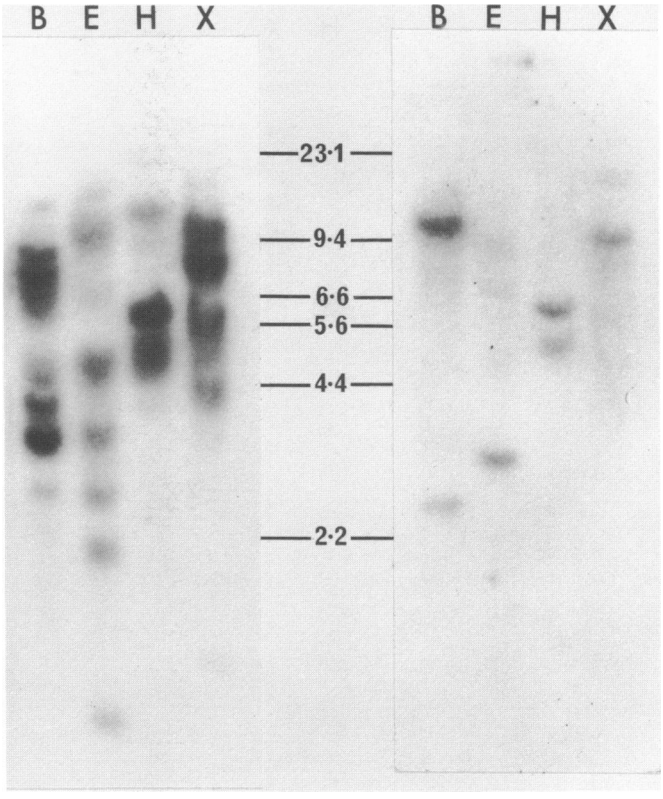

Figure 4. Restriction Fragments in Human Myeloid Cell DNA that hybridize to Human Gene Probes at High Stringency. The same blot as shown in Fig. 3 after a high stringency wash.

to be about twice the intensity of the majority of the variable region fragments. This fragment is identified by its absence from the patterns obtained using either the VκI probe or the subgroup-specific VκIV probe, which lacks the constant region sequence (See below). Under high stringency hybridization conditions (0.1 x SSC, 65°C), our cDNA probe detects, in addition to the C band, a single major V band and a small number of very minor V bands in the BgI II, Hind III and Xba I digests, as shown in Fig. 4. Two major V bands are seen in the EcoR I digests, each having about half the intensity of the single bands seen in the Bgl II, Hind III and Xba I digests. The simple patterns obtained with the cDNA probe differ from those obtained with the VκI probe in the high stringency wash, which largely reproduces the more complex array

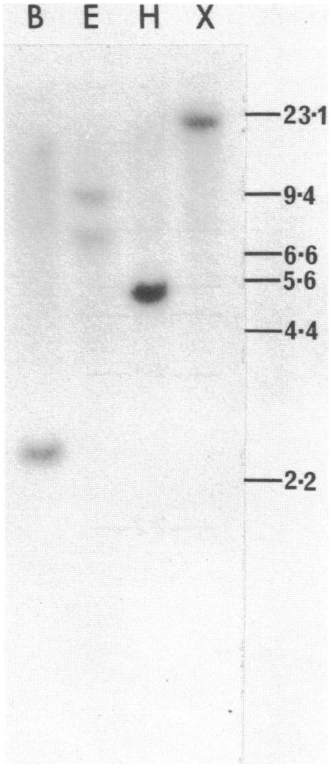

Figure 5. Restriction Fragments in Human Myeloid Cell DNA that hybridize to a Human IV Subgroup-Specific Probe at Low Stringency. A blot similar to the ones shown in Figs. 3 and 4 was hybridized with the subgroup specific probe described in the text.

of bands detected in the low stringency wash (Fig. 3).

A 156 base pair Hinf I fragment of pBl7κIV spanning codons 9-56 (Fig. la) was subcloned to provide a subgroup-specific probe. In contrast to the full length cDNA probe, the subgroup-specific probe lacks FR3, which is likely to crosshybridise to VκI and VκIII, and also Jκ and Cκ sequences. Only a single VκIV hybridizing band is seen in Bgl II, Hind III and Xba I digests of K562 DNA using this probe under low stringency washing conditions, as shown in Fig. 5. Two VκIV bands of lower intensity are again seen in the digests of the K562 DNA with EcoR I (cf. Fig.4). The digests of six other human DNA samples, however, revealed only a single V band, corresponding to the larger band of 11 kb (results not shown) in the K562 digest. The 7 kb band in the K562

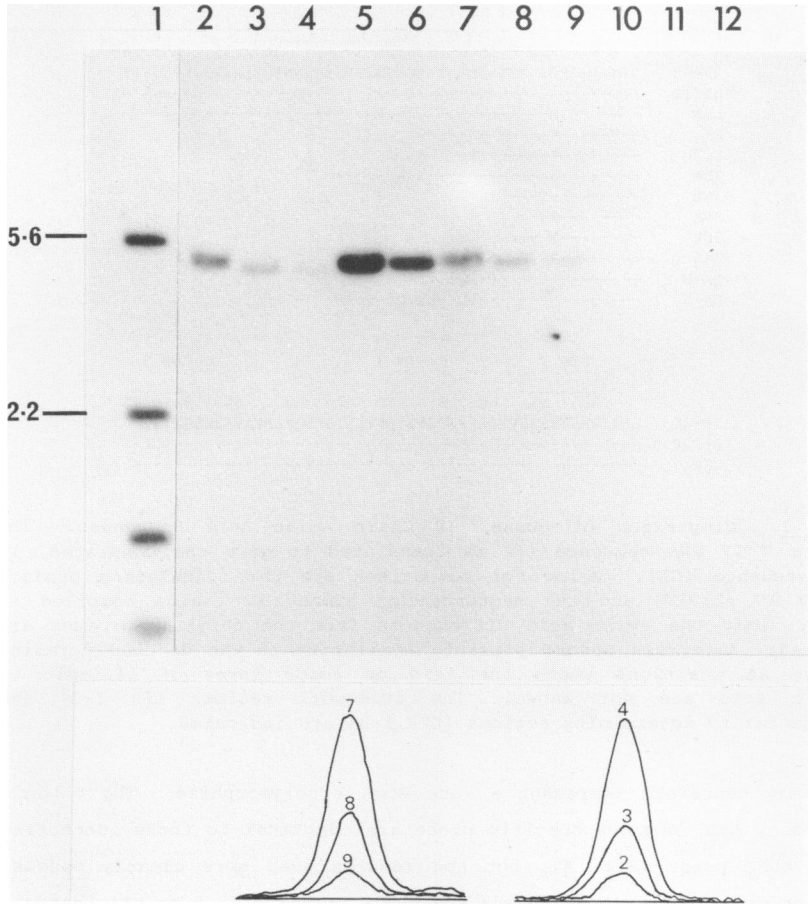

Figure 6. Estimation of the Haploid Genome Content of V IV Genes. In the upper panel lanes 2-4 show hybridization of the V IV specific probe respectively to 20, 10 and 5 µg of human peripheral blood DNA digested with Hind III. Lanes 5-9 contain cloned V IV standard, the 5.46 kb BamHI fragment from pJYMV.IV in the following amounts: lane 5 = 68 pg, lane 6 = 34 pg, lane 7 = 17 pg, lane 8 = 8.5 pg and lane 9 = 4.25 pg. A single copy gene equivalent in 10 µg human DNA is 16.8 pg using this standard. Samples in lanes 5-9 were mixed with 10 µg herring sperm DNA (Sigma) to control for variation in the transfer efficiency of a given fragment with the total amount of DNA. Lane 11 contains 10 µg herring sperm DNA and lane 12 contains 500 pg pJYM plus 10 µg herring sperm DNA. Neither of these components cause background hybridization. The lower panel shows densitometric traces of the autoradiogram reproduced above: the trace numbers refer to the gel lanes.

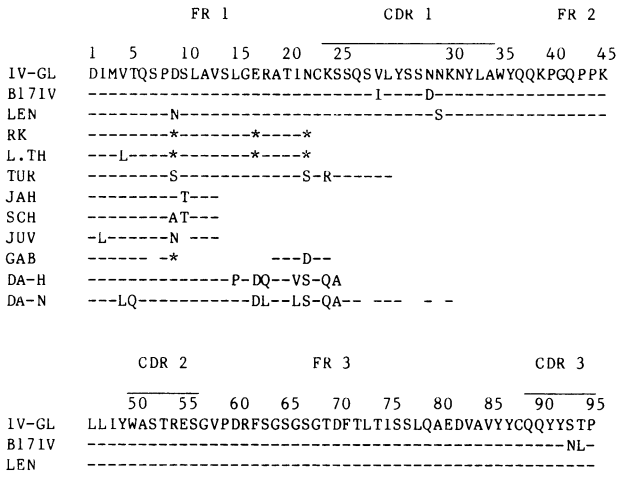

Figure 7. Comparison of Human IV Chain Amino Acid Sequences. The germline V IV DNA sequence (1) is translated to give the unmutated IV chain sequence (GL). Below for comparison are the translation product of pB17 IV (B171V) and the sequences of human IV chains compiled in Ref. 2. Only the amino acid differences from the germline product are indicated. Asterisks denote possible similarity to the deduced germline sequence at positions where the acid or amide forms of glutamic or aspartic acid are not known. The framework regions (FR 1-3) and complementarity determining regions (CDR 1-3) are indicated.

digest may therefore represent a rare EcoR I polymorphism. The V bands detected by the subgroup-specific probe are identical to those identified by the cDNA probe (Fig. 4), but the results seen more clearly because the minor crosshybridizing V bands and the C bands have been eliminated.

To estimate the number of genes represented by the single VκIV hybridizing bands in the restriction digests of genomic DNA, we measured the intensity of the signal relative to the intensities produced by known amounts of the sequence in cloned DNA. Since the 5 kb Hind III fragments (Fig. 5, Lane H) is in the size range for efficient transfer to nitrocellulose, we arranged for the cloned VκIV titration standards to be of similar size. Accordingly, the 156 bp HinFI fragment shown in Fig. 1 was sub-cloned into the largest NcoI fragment of pJYM (24) to generate pJYMVκIV. On BamH I digestion this recombinant yielded a 5.46 kb fragment containing the VκIV sequence, which was gel purified and used for calibration. Assuming the haploid human genome to contain 2.9×10^9 bp (25), 16.8 pg represents a single copy gene equivalent per 10 µg of human DNA. Standards representing from 0.25 to 4 genes

per 10 μg were subjected to electrophoresis alongside 5, 10 and 20 μg of peripheral blood DNA, digested with Hind III. The band intensities obtained on hybridization with the subgroup specific probe at low stringency were analyzed by densitometry as shown in Fig. 6. This titration reveals that there are 0.5 VκIV genes per haploid genome in the DNA sample. The closest integral number of genes is one. The weak signal of the VκIV band as compared with the unique C band in Figs. 4 and 6 may imply that the imperfect sequence identity between the expressed gene probe (owing to somatic mutations: see below) and the parent germline gene results in decreased stability of the hybrids.

In Fig. 7 we compare the translated VκIV sequence of pBl7κIV and the known V region sequences of human κ chains, obtained by amino acid sequencing, with the translated sequence of the germline VκIV gene (1). With the exceptions of the myelomas DA-N and DA-H, only a small number of amino acid substitutions are observed.

DISCUSSION

Our results, together with those of Klobeck *et al* in the accompanying paper (1), strongly point to the existence of a unique human VκIV gene. Our results are based upon the enumeration of fragments in the digests of human DNA with four restriction endonucleases and the determination of the quantity of hybridizing DNA in the case of the unique Hind III fragment using a κIV, subgroup-specific cloned cDNA probe. The conclusion of Klobeck *et al* is likewise based upon the enumeration of restriction fragments using, however, a subgroup-specific cloned genomic DNA probe. The differently derived VκIV probes gave consistent results in the Eco RI and Hind III digestions, common to both studies, revealing fragments of 11 kb and 5 kb, respectively. Our gene titration data add confidence to the conclusion based upon the restriction patterns: as pointed out by Klobeck *et al* (1), the duplication of an extensive length of human DNA, containing the majority of V genes (11), might lead to the occurrence of pairs of restriction fragments having the same electrophoretic mobility. However, VκIV does not appear to be present at the duplicated locus (11), and the titration here reveals that there is only one gene per haploid genome. The existence of a single VκIV gene simplifies the interpretation of sequence data in terms of the operation of different somatic processes in the generation of diversity.

One such process occurs when the V gene recombines with one of four

Jκ segments to form a functional κ chain gene (26). The identity of the Jκ segments coding for the 3' region of the B17κIV chain can be deduced from the DNA sequence (Fig. 1b); it is Jκ1, and codon 95 is composed of two residues from the VκIV segment and one 5' to the Jκ1 segment. We note that Jκ4 has been selected in the J1 lymphoma (1) and Jκ2 in the myeloma LEN (Ref.2, See below). Thus VκIV can recombine with at least three of the four Jκ segments (22). Four base substitutions in the germline Vκ IV segment near the junction with Jκ1 are observed in codons 92 and 93 of pB17κIV. A similar clustering of mutations in codons 92, 93 and 94 is seen in the rearranged Vκ IV gene isolated from the lymphoma J1 (1). This clustering of mutations around the point of recombinations between VκIV and Jκ1 raises the possibility that all are associated with the joining event, rather than occuring subsequently in the maturation of the immune response (27). Observation in immunoglobulin genes of joining associated mutations has led to models involving deletion (28) and insertion (29) of DNA.

Point mutations have occurred upstream of the joining region in the expressed VκIV gene sequence in the B17 lymphoblastoid cell line, as judged by the cDNA sequence. These occur at -4 in the leader peptide (a silent mutation from GGT to GGC), in CDR1 at codon 27B (changing the Ser codon, AGT, to a Val codon, AGC) and at codon 28 (changing the Asn codon, AAC, to an Asp codon, GAC), and at codon 50 in CDR 2 (a silent mutation of TGG to TGC). The clustered mutations in CDR3, which may have a different origin, as mentioned above, result in a change from AGTACT to AATTTA (codons 92 and 93) and substitution of Asn-Leu for Ser-Thr. These changes account for the four substitutions in the B17κIV chain relative to the putative germline counterpart (1). The non-silent mutations are all in CDR regions, two in CDR 1 and two in CDR 3, and may be related to the generation of an antibody κ chain with the anti-Streptococcus A specificity. In the lymphoma J1κ IV chain gene, the only mutation outside CDR 3 is a silent mutation at codon 46 in FR2 (1).

The amino acid changes from the translated germline gene sequence in the majority of sequenced VκIV chains (Fig. 7) can be accounted for by a small minimum number (between one and four) base substitutions in the germline VκIV gene. The myeloma amino terminal sequences for DA-N and DA-H may represent exceptions, to account for 8 out of 32 amino acid substitutions. There are two alternative explanations for the extent of divergence of the DA-N and DA-H sequences from the other κIV sequences.

They may be derived from a different germline gene or, may represent extreme somatic variants of the VκIV gene. Since it has been shown here and in the accompanying paper (1) that there is only a single VκIV gene, we favor the latter interpretation. Similar extents of somatic mutation have previously been observed (30).

In summary, we conclude that human VκIV chains are represented by a unique gene that is diversified exclusively by somatic processes. The proportion (0.05) of human κIV chains in the compilation of sequenced human κ chains (in Ref. 2) is about the same as the proportion (0.05) of VκIV genes in the total Vκ gene repertoire (9,10). The restricted gene population may therefore account for the restricted abundance of its products.

ACKNOWLEDGEMENTS

This work was supported by a Project Grant from the British Cancer Research Campaign. We thank Professor Klaus Eichmann for the B17X2 cell line, Dr Christine Polke for helpful discussions during the course of the work, Dr Terry Rabbitts for the VκI probe, and Prof. H.G. Zachau for the communication of the results in the accompanying paper prior to publication.

REFERENCES

1. Klobeck, H.G., Bornkamm, G.W., Combriato, G., Mocikat, R., Pohlenz, H.D. and Zachau, H.G. (1985) accompanying paper.
2. Kabat, E.A., Wu, T.T., Bilofsky, H., Reid-Muller, M., and Perry, H. Sequences of Proteins of Immunological Interest (National Institutes of Health, Bethesda 1983).
3. Nisonoff, A., Hopper, J.E., Spring, S.B. The Antibody Molecule (Academic Press 1975).
4. Wang, A.C., Fudenberg, H.H., Wells, J.V., and Roelcke, D. (1973) Nature New Biol. 243, 126-127.
5. Steinitz, M., Seppala, I., Eichmann, K., and Klein, G. (1979) Immunobiol. 156, 41-47.
6. Polke, C., Greger, B., Steinitz, M., and Eichmann, K. (1982) Immunobiol. 163, 95-109.
7. Molgaard, H.V., Wier, L., Kenten, J., Cramer, F., Klukas, C., Gould, H.J., and Birch, J.R. (1981) Biochemistry 20, 4467-4477.
8. Wier, L. (1982) Ph.D. Thesis Univ. Lond.
9. Bentley, D.L. and Rabbitts, T.H. (1981) Cell 24, 613-623.
10. Klobeck, H.G., Solomon, A. and Zachau, H.G. (1984) Nature 309, 73-75.
11. Pech, M., Smola, H., Pohlenz, H.-G., Straubinger, B., Gerl, R., and Zachau, H.G. (1985) J.Mol.Biol. 183, 291-299.
12. Land, H., Grez, M., Hauser, H., Lindenmaier, W., and Schutz, G. (1981) Nucl. Acids. Res. 2252-2266.
13. Gough, N.M., Webb, E.A., Cory, S., and Adams, J.M. (1980) Biochemistry 19, 2702-2710.

14. Twigg, A.J. and Sherratt, D. (1980) Nature 283, 216-218.
15. Smith, H.O. (1980) Methods Enzymol. 65, 371-380.
16. Hanahan, D. (1983) J.Mol.Biol. 166, 557-580.
17. Grunstein, M. and Hogness, D. (1975) Proc. Natl. Acad. Sci. 72, 3961-3965.
18. Amersham International (1980) Technical Bulletin TB80/3, Amersham, UK.
19. Maxam, A.M. and Gilbert, W. (1980) Methods Enzymol. 65, 499-560.
20. Maniatis, T., Fritsch, E.F. and Sambrook, J. (1982) Molecular Cloning A Laboratory Manual (Cold Spring Harbour Laboratory, New York).
21. Challberg, M.D. and Englund, P.T. (1980) Methods Enzymol. 65, 39-43.
22. Hieter, P.A., Maizel, J.V. Jnr. and Leder, P. (1982) J. Biol. Chem. 257, 1516-1522.
23. Lozzio, C.B. and Lozzio, B.B. (1975) Blood 45, 321-334.
24. Lusky, M. and Botchan, M. (1981) Nature 293, 79-81.
25. Kornberg, A. (1980) DNA Replication. (San Francisco: W.H. Freeman and Co.)
26. Tonegawa, S. (1983) Nature 302, 575-581.
27. Griffiths, G.M., Berek, C., Kaartinen, M. and Milstein, C. (1984) Nature 312, 271-275.
28. Karjalainen, K. and Coleclough, C. (1985) Nature 314, 544-546.
29. Alt, F. and Baltimore, D. (1982) Proc. Natl. Acad. Sci. 79, 4118-4122.
30. McKean, D., Huppi, K., Bell, M., Staudt, L., Gerhard, W. and Weigert, M. (1984) Proc. Natl. Acad. Sci. 81, 3180-3184.